

xtbreak: Testing for structural breaks in Stata

German Stata Conference 2021

Jan Ditzen¹, Yiannis Karavias², Joakim Westerlund³

¹Free University of Bozen-Bolzano, Bozen, Italy
www.jan.ditzen.net, jan.ditzen@unibz.it

²University of Birmingham, UK
<https://sites.google.com/site/yianniskaravias/>
i.karavias@bham.ac.uk

³Lund University, Lund, Sweden

June 25, 2021

Motivation

- In time series or panel time series structural breaks (or change points) in the relationships between key variables can occur.
- Estimations and forecasts depend on knowledge about structural breaks.
- Structural breaks might influence interpretations and policy recommendations.
- Break can be unknown or known and single and multiple breaks can occur.
- Examples: Financial Crisis, oil price shock, Brexit Referendum, COVID19,...
- Question: Can we estimate when the breaks occur and test them?

Literature

- Time Series:
 - ▶ Andrews (1993) test for parameter instability and structure change with unknown change point.
 - ▶ Bai and Perron (1998) propose three tests for and estimation of multiple change points.
- Panel (Time) Series:
 - ▶ Wachter and Tzavalis (2012) single structural break in dynamic independent panels.
 - ▶ Antoch et al. (2019); Hidalgo and Schafgans (2017) single structural break in dependent panel data.
 - ▶ Ditzen et al. (2021); Karavias et al. (2021) single and multiple breaks in panel data with cross-section dependence.
- `xtbreak` introduces tests for multiple structural breaks in time series and panel data based on Bai and Perron (1998) and Ditzen et al. (2021); Karavias et al. (2021).

Econometric Model I

- Static linear panel regression model with s breaks:

$$\begin{aligned}y_{i,t} &= x'_{i,t}\beta + w'_{i,t}\delta_1 + u_{i,t}, & t = 1, \dots, T_1, \quad i = 1, \dots, N \\y_{i,t} &= x'_{i,t}\beta + w'_{i,t}\delta_2 + u_{i,t}, & t = T_1 + 1, \dots, T_2 \\&\dots \\y_{i,t} &= x'_{i,t}\beta + w'_{i,t}\delta_{s+1} + u_{i,t}, & t = T_s, \dots, T\end{aligned}$$

- $\tau_s = (T_1, T_2, \dots, T_s)$ are break points of the s breaks.
- x_t is a $(1 \times p)$ vector of variables without structural breaks.
- w_t is a $(1 \times q)$ vector of variables with structural breaks.
- Fixed effects can be included in $x_{i,t}$, pooled constant can be included in $x_{i,t}$ or $w_{i,t}$
- Error $u_{i,t}$ contains unobserved heterogeneity ($u_{i,t} = f'_t \gamma_i \epsilon_{i,t}$).

Econometric Model II

- The model can be expressed in matrix form:

$$Y_i = X_i\beta + Z_i(\tau_s)\delta + U_i \quad (1)$$

- where $Y_i = (y_{i,1}, \dots, y_{i,T})'$, $W_i = (w_{i,1}, \dots, w_{i,T})'$, $\delta = (\delta'_1, \dots, \delta'_{s+1})'$ and:

$$W_i(\tau_s) = \begin{pmatrix} w_{1,i} & 0 & \dots & 0 \\ 0 & w_{2,i} & \dots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & \dots & \dots & w_{s+1,i} \end{pmatrix}$$

- $w_{s,i}$ is $(T_s \times q)$.
- Aim: Test if and when breaks occur.

Hypotheses

- Three hypotheses (Bai and Perron, 1998):
 - 1 No break vs. s breaks
 $H_0 : \delta_1 = \delta_2 = \dots = \delta_{s+1}$ vs $H_1 : \delta_k \neq \delta_j$ for some $j \neq k$.
 - 2 No break vs $1 \leq s \leq s^*$ breaks
 $H_0 : \delta_1 = \delta_2 = \dots = \delta_{s+1}$ vs $H_1 : \delta_k \neq \delta_j$ for some $j \neq k$ and $s = 1, \dots, s^*$
 - 3 s breaks vs $s + 1$ breaks
 $H_0 : \delta_j = \delta_{j+1}$ for one $j = 1, \dots, s$ vs. $H_1 : \delta_j \neq \delta_{j+1}$ for all $j = 1, \dots, s$.
- Next question: know or unknown breakpoints?

Tests

Unknown Breakpoints

- Main idea: if the model has the true number of breaks and the true point in time, then the SSR should be smaller than for a model with a larger or smaller number of breaks.
- No further knowledge of the break points required.

Test Hypothesis 1 I

No break vs. s breaks

$$H_0 : \delta_1 = \delta_2 = \dots = \delta_{s+1} \text{ vs } H_1 : \delta_k \neq \delta_j \text{ for some } j \neq k$$

- Wald test with test statistic:

$$F_T(\tau_s^0) = \frac{N(T - p - (s + 1)q) - p - (s + 1)q}{sq} \hat{\delta}' R' \left(R \hat{V}(\hat{\delta}) R' \right)^{-1} R \hat{\delta}$$

- R imposes the restrictions such that $R\delta' = (\delta'_1 - \delta'_2, \dots, \delta'_s - \delta'_{s+1})'$.
- $\hat{V}(\hat{\delta})$ is an estimate of the variance.

Test Hypothesis 1 II

No break vs. s breaks

- If the break dates are known, then (Andrews, 1993)

$$F_T(\tau) \sim \chi^2(sq).$$

- If the break dates are unknown, then $supF$ test statistic is used:

$$\sup F_T(s, q) = \sup_{\tau \in \tau_\eta} F_T(\tau, q)$$

- τ_ϵ is a subset of $[0, T]^s$ and represent all possible combination of break points with a minimal length of each set of η .
- Asymptotic critical values depending on the number of breaks s and regressors q are given in Bai and Perron (1998, Table 1).

Test Hypothesis 2 I

No break vs. $1 \leq s \leq s^*$ breaks

- Test if a maximum of s^* breaks occurs.
- "Double Maximum" test, where the maximum of the test using hypothesis 1 for the number of breaks between 1 and s^* is taken.

$$\text{WDmax}F_T(s, q) = \max_{1 \leq s \leq s^*} \left\{ \frac{c_{\alpha, 1, q}}{c_{\alpha, s, q}} \sup_{\tau \in \mathcal{T}_\eta} F_T(\tau, q) \right\}$$

- $c_{\alpha, s, q}$ is the critical value at a level of α for s breaks and q regressors.
- Asymptotic critical values depending on the number of breaks s and regressors q are given in Bai and Perron (1998, Table 1).

Test Hypothesis 3 I

s breaks vs. $s + 1$ breaks

- Idea: test each s segments for an additional break within the segment.

$$F(s + 1|s) = \frac{SSR(\hat{T}_1, \dots, \hat{T}_s)}{\min_{1 \leq j \leq s+1} \left\{ \inf_{\tau \in \Lambda_{j,\eta}} SSR(\hat{T}_1, \dots, \hat{T}_{j-1}, \tau, \hat{T}_j, \dots, \hat{T}_s) \right\}} \hat{\sigma}_s^2$$

$$\Lambda_{j,\eta} = \left\{ \tau; \hat{T}_{j-1} + (\hat{T}_j - \hat{T}_{j-1}) \eta \leq \tau \leq \hat{T}_j - (\hat{T}_j - \hat{T}_{j-1}) \eta \right\}$$

$$\hat{\sigma}_s^2 = \frac{SSR(\hat{T}_1, \dots, \hat{T}_s)}{N(T - 1) - sq - p}$$

$$SSR(\hat{T}_1, \dots, \hat{T}_{s+1}) = \min_{\tau \in \mathcal{T}_\eta} SSR(\tau)$$

Test Hypothesis 3 II

s breaks vs. $s + 1$ breaks

- Looks complicated.... but it is essentially the difference of the minimum of combinations of the SSR with s and $s + 1$ breaks.
- Asymptotic critical values depending on the number of breaks s and regressors q are given in Bai and Perron (1998, Table 2).

Finding unknown breaks

- Tests are based on minimising the SSR, but how to find the minimal SSR?
- `xtbreak` implements the dynamic programming algorithm from Bai and Perron (2003). Idea is to calculate the SSR for all *necessary* subsamples.
- For example: Break in period 2 ($T_1 = 2$), then $SSR = SSR(1, 2) + SSR(2, T)$.

		1	2	End 3	...	T
Start	1	•	$SSR(1, 2)$	$SSR(1, 3)$...	$SSR(1, T)$
	2		•	$SSR(2, 3)$...	$SSR(2, T)$
	3			•		$SSR(3, T)$
	⋮				•	
	T					•

xtbreak¹

```
xtbreak test depvar [indepvars] [if] [, break_point_options  
panel_options nobreakvariables(varlist ts) noconstant  
breakconstant vce(ssr|hac|nw) ]
```

If the breakpoint is known then `break_point_options` are:

```
breakpoints(numlist [ ,index ])
```

If the breakpoint is unknown then `break_point_options` are:

```
hypothesis(1|2|3) breaks(real) minlength(real) level(real)  
error(real) wdmax
```

- ▶ `breaks(real)` sets the number of breaks.
- ▶ `breakpoints(numlist)` sets the breakpoints.
- ▶ `vce` is the variance/covariance estimator.

¹This command is work in progress. Options, functions and results might change.

xtbreak

panel_options are specific for panel data sets:

```
nofixedeffects csd csa(varlist, deterministic[(varlist)])  
csanobreak(varlist, deterministic[(varlist)])
```

- ▶ `nofixedeffects` omits fixed effects model. If `noconstant` not used, assume pooled OLS model.
- ▶ `csa` and `csanobreak` define variables added as cross-section averages. Suboption `deterministic` treats variables as deterministic cross-section averages.
- ▶ `csd` automatically select cross-section averages.

xtbreak update

- Updates `xtbreak` from [GitHub](#).

Example

Simulated Data

- We generate a panel dataset with 100 cross-sectional units ($N = 100$) and quarterly data from 1990Q1 to 2002Q4 ($T = 52$).
- Assume two breaks in variable w and no breaks in variable x .
- The DGP is:

$$y_{i,t} = \alpha_i + \beta_0 + \gamma_s w_{i,t,s} + \beta x_{i,t} + \epsilon_{i,t}$$

- with $\gamma_1 = 2, \gamma_2 = 1, \gamma_3 = 0.5, \beta = 1$ and $\epsilon_{i,t} \sim N(0, 10)$.

Simulated Data

How does the data look like?

- Definition of a structural break: Period when the coefficient changes, i.e. T_s , or last period with a given coefficient.
- In the example the break occurs in period 1994Q1 and 1998Q2, or after 17 respectively 34 periods.

```
. list id t b if id == 1 & ///  
> ((t > tq(1993q3) & t < tq(1994q4)) | ///  
> (t > tq(1997q4) & t < tq(1999q1)))
```

	id	t	b
16.	1	1993q4	2
17.	1	1994q1	2
18.	1	1994q2	1
19.	1	1994q3	1
33.	1	1998q1	1
34.	1	1998q2	1
35.	1	1998q3	.5
36.	1	1998q4	.5

Known Breakdates I

Test for no vs 2 breaks

- Since we generated the data, we know the breaks.
- We want to test for no break against the alternative of a break in 1994Q1 and 1998Q2.
- Hypothesis: H_0 : no breaks vs H_1 : $s = 2$

```
. xtbreak test y w , nobreakvar(x) ///  
> breakpoints(tq(1994Q1) tq(1998q2))  
  
Test for multiple breaks at unknown breakdates  
(Ditzen, Karavias & Westerlund. 2021)  
H0: no breaks vs. H1: 2 break(s)  
  
W(tau) = 900.98  
p-value (F) = 0.00  
p-value (chi) = 0.00
```

- We reject the hypothesis that there is no break.

Known Breakdates II

Test for no vs 2 breaks

- Instead of the dates using `tq()` we can directly use the index of the time periods:

```
. xtbreak test y w , nobreakvar(x) ///  
> breakpoints(17 34, index)
```

Test for multiple breaks at unknown breakdates
(Ditzen, Karavias & Westerlund. 2021)

H0: no breaks vs. H1: 2 break(s)

```
W(tau)    =    900.98  
p-value (F) =      0.00  
p-value (chi) =      0.00
```

- Result is the same.

Unknown Breakdates

Test for no vs 2 breaks

- We can test the same hypothesis but with unknown breaks.
- Hypothesis: H_0 : no breaks vs H_1 : $s = 2$ (2 breaks)

```
. xtbreak test y w , nobreakvar(x) ///
> breaks(2) hypothesis(1)

Test for multiple breaks at unknown breakdates
(Ditzen, Karavias & Westerlund. 2021)
H0: no break(s) vs. H1: 2 break(s)
```

	Bai & Perron Critical Values			
Test Statistic	1% Critical Value	5% Critical Value	10% Critical Value	
supW(tau)	878.65	9.36	7.22	6.28

Estimated break points: 1994q1 1997q4

- We reject the hypothesis of no break.
- The estimated break date for the first break is correctly estimated, the estimate for the second break is mildly off.

Unknown Breakdates

Test of no vs up to 2 breaks

- We can test if the number of breaks is up to or smaller than a given number.
- Assumptions that we have at most 2 breaks. The hypothesis is: H_0 : no breaks vs H_1 : $1 \leq s \leq 2$ breaks.

```
. xtbreak test y w , nobreakvar(x) ///
> breaks(2) hypothesis(2)
```

Test for multiple breaks at unknown breakdates
(Ditzen, Karavias & Westerlund. 2021)
H0: no break(s) vs. H1: $1 \leq s \leq 2$ break(s)

	Test Statistic	Bai & Perron 1% Critical Value	Critical Values 5% Critical Value	10% Critical Value
UDmax(tau)	1613.92	12.37	8.88	7.46

Estimated break points: 1994q1
* evaluated at a level of 0.95.

Unknown Breakdates

Test of s against $s+1$ breaks

- So far under the null there was no break.
- Hypothesis 3 allows to test for s breaks under the null and $s + 1$ breaks under the alternative.
- Hypothesis in the example: $H_0 : 1$ break vs $H_1 : 2$ breaks.

```
. xtbreak test y w , nobreakvar(x) ///
> breaks(2) hypothesis(3)
```

Test for multiple breaks at unknown breakpoints
(Ditzen, Karavias & Westerlund. 2021)
H0: 1 vs. H1: 2 break(s)

	Test Statistic	Bai & Perron Critical Values		
		1% Critical Value	5% Critical Value	10% Critical Value
F(s+1 s)*	0.42	13.89	10.13	8.51

* s = 1

Conclusion

- Introduced new community contributed package called `xtbreak`
- Test for breaks at known and unknown points in time.
- Three tests for time series and panel data included, following Bai and Perron (1998); Ditzen et al. (2021); Karavias et al. (2021).
- For the ado files, further details and examples see our [GitHub](#) page or

```
net install xtbreak, from(https://janditzen.github.io/xtbreak/)
```

- What's next:
 - ▶ Work in progress!
 - ▶ Confidence intervals for estimated breaks with `xtbreak estimate`.
 - ▶ Variance/Covariance estimator.
 - ▶ Improve speed.

References I

- Andrews, D. W. K. 1993. Tests for Parameter Instability and Structural Change With Unknown Change Point. Econometrica 61(4): 821–856.
- Antoch, J., J. Hanousek, L. Horvath, M. Huskova, and S. Wang. 2019. Structural breaks in panel data: Large number of panels and short length time series. Econometric Reviews 38(7).
- Bai, B. Y. J., and P. Perron. 1998. Estimating and Testing Linear Models with Multiple Structural Changes. Econometrica, 66(1): 47–78.
- Bai, J., and P. Perron. 2003. Computation and analysis of multiple structural change models. Journal of Applied Econometrics 18(1): 1–22.
- Ditzen, J., Y. Karavias, and J. Westerlund. 2021. Testing for Multiple Structural Breaks in Panel Data .
- Hidalgo, J., and M. Schafgans. 2017. Inference and testing breaks in large dynamic panels with strong cross sectional dependence. Journal of Econometrics 96(2).

References II

- Karavias, Y., J. Westerlund, and P. Narayan. 2021. Structural Breaks in Interactive Effects Panels and the Stock Market Reaction to COVID-19 .
- Wachter, S. D., and E. Tzavalis. 2012. Detection of structural breaks in linear dynamic panel data models. Computational Statistics & Data Analysis .